

# CWIC Team OpenSearch Action Items

Replaces GoogleDocs <https://docs.google.com/document/d/1Hq4Hi7OFY13stCgVL1XSsLGhi9D4enlhfBTOHpybtsQ/edit#>

**Details about many of the action items below, with example code and scenarios were initially captured at:**

[https://docs.google.com/document/d/1rfX1YXnlI66DpX0GOnYtUQ4FK\\_eiPncN4WGrdRME2U/edit](https://docs.google.com/document/d/1rfX1YXnlI66DpX0GOnYtUQ4FK_eiPncN4WGrdRME2U/edit)

## Completed Action Items

- ☐ 3. CWIC, GCMD, ECHO - Use parameter names consistently (searchTerms, count, startIndex, startPage) - definitions? - see Calin note below  
Status >> In process
  - ☐ ECHO Doug Newman says 'You want us to change our parameter names even though they are linked to the canonical terms in the OSDD? We don't want to do that. The risk on breaking clients who are not getting our OSDD dynamically is too great. The url template parameter supports this via pw={startPage?} where we can assign any parameter name to map to the os:startPage concept. This allows flexibility on the server side and is "a good thing"

searchTerms: value must be URL encoded  
count: number of results per page  
startIndex: index of the first search result  
startPage: page number of the search results  
language: string indicating language  
inputEncoding: default "UTF-8"  
outputEncoding: default "UTF-8"

Doug Newman we support searchTerms, count and startIndex (we need to switch that to startPage - see other action). I don't see the need to support language, inputEncoding and outputEncoding. The spec does not detail whether they should be mandatory or not.

Calin Duma I am asking to add support for parameter names mentioned in the opensearch standard document. You can add additional parameter names without breaking compatibility with the old parameter names if you wish to continue supporting existing clients. The parameter names are mentioned in the OpenSearch standard document. They are also listed below for your convenience:

- ☐ 4. CWIC, GCMD - Align namespace declarations  
Status >> In process, awaiting final version of GCMD response
- ☐ 5. GCMD - inconsistency in startIndex (0 indexed? 1 indexed?)  
Status >> In process - startIndex fails on 0 (bug) and defaults to 1 if not provided. Is this what you want?
- ☐ 6. ECHO - namespace declarations in Atom response? Doug Newman says 'Is this action related to moving these declarations from the entry to the feed? If so, this is scheduled for the 10.73 release'
- ☐ 7. CWIC - namespace declaration for datasetId, granuleId attributes (CWIC namespace URL?)  
Status >> Still awaiting analysis.
- ☐ 19. CWIC, GCMD - is startIndex 0-indexed or 1-indexed?  
Status >> pending - Archie: I believe it should be 1-indexed.
- ☐ 20. CWIC, GCMD - requesting startPage=startPage+1 should retrieve next count results?  
Status >> pending
- ☐ 21. GCMD - rss does not include aoi, toi  
Status >> pending - Do you have a spec for this? Also, CWIC is not interested in RSS is it?  
Calin Duma This is exactly why I thought having a summary of the action items extracted from the original document is not a good idea. Many of the questions in this document are answered, with code, example and more details in the original doc. Your fragment of interest is:

Q8: This is not a big deal, the rss output doesn't have the aoi and toi info, we won't use rss though. The sample request below is from your notes to Shao's document:

<http://gcmd.gsfc.nasa.gov/KeywordSearch/OpenSearch.do?searchTerms=water&output=rss&count=3&startIndex=1&startPage=1&geoBox=-150,-45,-150,45&timeStart=2000-02-14&timeEnd=2010-02-14&Portal=GCMD>

- ☐ 24. GCMD - georss response correct for multiple bounding boxes, and must be a single bounding box  
Status >> pending
- ☐ 25. GCMD - how do temporal searches work?  
Status >> pending - update: adding better support for open ended searches, found that open search required both start and end where as backend code supported both or supplying only a start. No reason to not support only having an end date, so

experimenting with this. If a start date is not given, use the minimum date provided by language. If no end date, uses max date. Code then does an intersection look up from lucene.

- ☐ 27. CWIC - pagination function aligned with GCMD?  
Status - in process
- ☐ 29. ECHO - format of georss response should be as described by the spec [Doug Newman](#) says "This is detailed in the [Open Search Geo extension](#)"
- ☐ 30. ECHO - explicit namespace declarations [Doug Newman](#) says 'please elaborate? Is this feed versus entry level? If so, see above'
- ☐ 31. CWIC - ensure opensearch URL element points consistently to the appropriate instance (OSDD and results)
- ☐ 33. CWIC - check namespace declarations for attributes in the Query element (datasetId)
- ☐ 34. ECHO - we appear to be using startIndex as a cursor argument in ECHO Open Search. That would imply that several other orgs are making the same assumption otherwise we would have had a slew of bug reports. We need to discuss this. **This will be in 10.73**

**Calin Duma** This is one area where Archie and I had a different interpretation of the standard (probably mine was influenced by the way the GCMD opensearch worked, which I understood after some experimentation and made sense to me) and we needed some feedback from you to help us figure out the correct interpretation. Please see Q2 in the original google doc: [https://docs.google.com/document/d/1rfX1YXnLI66DpX0GOnYtUQ4FK\\_eiPncN4WGrdRME2U/edit](https://docs.google.com/document/d/1rfX1YXnLI66DpX0GOnYtUQ4FK_eiPncN4WGrdRME2U/edit)

From Calin:

*The param names are captured correctly from the google doc, complete list below:*  
*searchTerms: value must be URL encoded*  
*count: number of results per page*  
*startIndex: index of the first search result*  
*startPage: page number of the search results*  
*language: string indicating language*  
*inputEncoding: default "UTF-8"*  
*outputEncoding: default "UTF-8"*

Not all data providers support the notion of paging. Paging is trivially implemented from only a starting record offset from the beginning of the result set and a count of items to retrieve. What additional functionality does startPage provide, and should it be mandatory?

**Calin Duma** I realize that data providers might not support it directly, I was however asking for CWIC opensearch to support it directly.

The preferred method for me as a client was startPage + count (as stated in the original document Q2) since really that is what I am displaying in my pagination code (pages and not result offset + count). However, if CWIC and GCMD decide to go with startIndex, I can support that without any problems.

(Archie) Here is a simple thought experiment. Your query finds N matching records at the remote system, and they happen to appear in the result set in order - record #1 first, record #2 second, ... record #N last. You now ask for records with startIndex=5, startPage=20 and count=2.

**Calin Duma** I would never ask for that, I would ask for startPage from 1 to the last page based on count. That is exactly what I am doing now with GCMD, I am always using the same value for startIndex since I am not interested in iterating in a page (I had a very similar example in Q2).

**Archie Warnock** What is the record number of the first record the server returns to you, and where in the "page" that is returned does the requested record (with startIndex=5) appear? Remember - the remote system may have no notion of page numbers at all. I ask this because it still seems to me that either startIndex and startPage are contradictory, or one of them is irrelevant. Once we resolve this, I think we'll all understand the use of the parameters.

**Calin Duma** That was exactly my point in Q2. What I was asking in the document and some of the email exchanges that we had is that both CWIC and GCMD to support the same paging mechanism (my preferred approach would not make use of startIndex at all since it is confusing). If Doug mentions that startIndex is an offset in the result set and not in a page (which is what Archie understood from the OpenSearch standard and I can see how that is also possible in reading the document) and GCMD and CWIC agree to use the same pagination approach, I can easily change my code if you decide on the approach that is not my preferred one. The important part is captured in the CWIC Q7 (which refers to GCMD Q7 for more GCMD implementation details / issues):

**Q7:** GCMD and CWIC should support identical opensearch result pagination. Please see GCMD Q7 above. The current approach is not consistent between GCMD and CWIC.

**Calin Duma** One additional point that made me think that the GCMD implementation is what the standard document conveys. We have two possibilities that we are trying to choose from:

1. startIndex is to be used for iterating inside the result set (and not inside a page) - this was Archie's interpretation. However, if both startIndex and startPage are present and Archie's interpretation is applied, the request can be ambiguous.
2. startIndex is to be used for establishing the start offset inside the current page (this is my interpretation and the way GCMD tried to implement it, there are some issues to be solved still in their implementation but the reasoning remains the same). This would result in a clear pagination approach with no request ambiguities. It doesn't mean this is the right interpretation though and that is what we needed Doug to chime in. [Doug Newman](#) says ECHO will now replace the incorrect use of startIndex with startPage. We do not support startIndex. I would suggest CWIC and IDN take the same approach. I would like to make it a best practice.

- ☐ 35. CWIC, GCMD, ECHO - add next, prev and last parameters to Atom response? [Doug Newman](#) says ECHO does this for self, prev and next. Not sold on the utility of 'last' though.

[Team Action Items](#)

[CWIC Action Items](#)

[GCMD Action Items](#)

[ECHO Action Items](#)

Thomas Cherry Note on entry ID:

The atom feed in the next version of GCMD will filter out originating node when even it can be assumed by the software. In cases that it can not, GCMD provides the originating node and entry id in a GCMD namespace for parsing. Please review attached [atom.xml](#) file